ESD-TR-91-094

AD-A243 183

DTIC ELECTE DEC 4 1991 S C D

Technical Report 927

Performance Measures for Adaptive Decisioning Systems

R.Y. Levine
T.S. Khuon

11 September 1991

Lincoln Laboratory
MASSACHUSETTS INSTITUTE OF TECHNOLOGY
LEXINGTON, MASSACHUSETTS

Prepared for the Department of the Army
under Air Force Contract F19628-90-C-0002.

91-16959

This report is based on studies performed at Lincoln Laboratory, a center for research operated by Massachusetts Institute of Technology. The work was sponsored by the Department of the Army under Air Force Contract F19628-90-C-0002.

This technical report has been reviewed and is approved for publication.

FOR THE COMMANDER

Hugh L. Southall

Hugh L. Southall, Lt. Col., USAF
Chief, ESD Lincoln Laboratory Project Office

MASSACHUSETTS INSTITUTE OF TECHNOLOGY
LINCOLN LABORATORY

PERFORMANCE MEASURES FOR ADAPTIVE DECISIONING SYSTEMS

R.Y. LEVINE
T.S. KHUON
Group 93

TECHNICAL REPORT 927

11 SEPTEMBER 1991

LEXINGTON MASSACHUSETTS

ABSTRACT

Performance measures are derived for data-adaptive hypothesis testing by systems trained on stochastic data. The measures consist of the averaged performance of the systems over an ensemble of training sets. The uncertainties derivable from training sets represent an irreducible uncertainty inherent in the learning procedure. Data-adaptive system estimates are contrasted with classical hypothesis testing, in which optimum tests are based on an assumed data model. In addition, a performance estimate for the maximum *a posteriori* probability (MAP) N-hypothesis test is derived based on a neural-net formulation of the test. The performance of adaptive systems on a binary test of uniformly distributed data is compared with the data-adaptive and MAP estimates. The adaptive systems considered are linear extrapolation from data (LINEXT) and a back-propagation neural net (BPNN).

ACKNOWLEDGMENTS

We thank Mitch Eggers for the suggestion that neural nets are known to perform the MAP hypothesis test. The reading of the text by Sun Levine is very much appreciated.

TABLE OF CONTENTS

LIST OF ILLUSTRATIONS

1. INTRODUCTION

Hypothesis testing by a data-adaptive system is fundamentally different from classical hypothesis testing. In the former, a representative data set corresponding to known hypotheses is used to train the system. System parameters are varied until the system training set to hypothesis space mapping best approximates the known map. Two assumptions, a sufficiently representative training set and the ability of the system to associate, are required to extend the map to arbitrary data [1]. In contrast, classical hypothesis testing derives from an assumed model for the data, often a signal in Gaussian noise, from which optimum tests are defined [2].

In this report, performance measures are derived based only on the procedure by which an adaptive system is trained. We assume that if a system is perfectly trained on a representative data set for each hypothesis, an appropriate performance estimate is the averaged performance over the ensemble of training sets. This averaged performance, which is computed in terms of training-set size and data distributions, reflects an uncertainty inherent in the learning procedure.

A data-adaptive system of particular interest is the neural net. Relative to the now-conventional neural-net taxonomy [1,3], we will consider only the back-propagation mapping network [4–7]. This network adapts internal parameters toward the approximation of a functional mapping, which for hypothesis testing is the data input to hypothesis space output map. Alternative neural-net architectures, such as those employing Kohonen learning [8], attempt to store data distributions internally rather than to approximate a map to the hypothesis space. Neural-net classifiers generally perform as well as conventional techniques on a variety of problems including linear, Gaussian, and k-nearest-neighbor algorithms [3], [9–12]. Neural nets have also been configured to perform maximum *a posteriori* probability (MAP) [13] and maximum likelihood tests [14] for arbitrary input distributions.

In Section 3, training-set-based performance measures are derived for a data-adaptive system on an arbitrary data-based N-hypothesis test. A MAP test is also formulated and represented in a neural-net structure. A possible neural-net representation of the MAP test contains N output neurons (processing elements). For a net input x, the ith neuron outputs $p(H_i|x)\epsilon[0, 1]$, which is the conditional probability for hypothesis H_i, $i = 1, \ldots, N$. The training-set-based and MAP estimates are applied in Section 3 to a binary hypothesis test on uniformly distributed data. These measures are compared to the computed performance of adaptive systems such as a linear extrapolation from the training set and a back-propagation neural net. Linear extrapolation (LINEXT) simply chooses the hypothesis of the nearest neighbor to the input, whereas a back-propagation neural net (BPNN) is trained to minimize the summed difference between net outputs and targets over the training set [4]. Both tests are data-adaptive in that the algorithms are defined using a training set. Section 3 shows that systems trained to the exact training-set map most closely match the training-set-based estimates. These systems are contrasted with systems trained on data biases, which are better approximated by Bayesian performance.

2. ADAPTIVE-SYSTEM PERFORMANCE MEASURES

In this section two performance measures are defined for data-adaptive systems: the training-set-based estimate, in which the statistics of the training set determine the performance, and the MAP test estimate. The MAP hypothesis test is formulated with an assumed neural-net structure for the data input to hypothesis space output map.

2.1 Training-Set-Based Measures

In this subsection the performance of an adaptive system is approximated from the statistics of the training set. Consider the training of an adaptive system for the testing of hypotheses $H_1, \ldots, H_N$ with prior probabilities $p(H_i), i = 1, \ldots, N$. The prior probabilities are normalized to unity by the condition $\sum_{i=1}^{N} p(H_i) = 1$. The input to the system is the data value $x \epsilon \mathcal{R}$, which is obtained by the observation of stochastic phenomena reflecting the set of possible hypotheses. We denote the operation of observing the phenomena, OBS, from which the data value x is obtained. The OBS-generated data value x is input to the adaptive system, which has an output $\vec{u} = (u_1, \ldots, u_N)$, with u_j nonzero corresponding to hypothesis $H_j, j = 1, \ldots, N$. Figure 1 contains a schematic of the OBS and adaptive system operations.

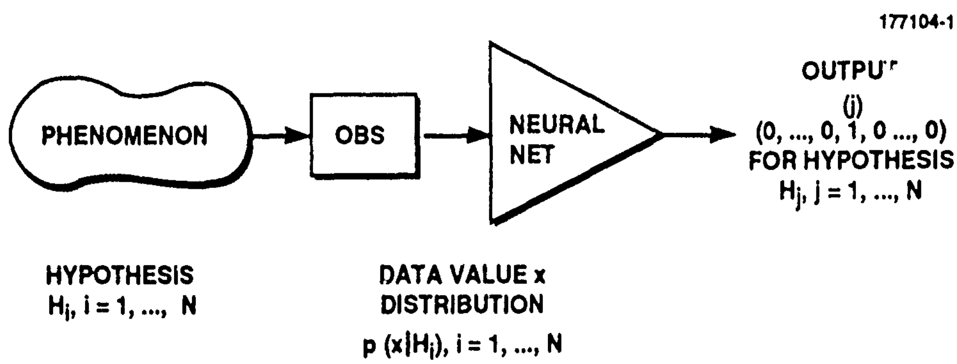

Figure 1. Schematic of the OBS and adaptive system operations. Hypotheses H_i, $i = 1, \ldots, N$, OBS output x, neural-net output $\vec{u}$.

The data value x is assumed to have a conditional probability distribution $p(x|H_i), i = 1,\ldots,N$, with hypothesis H_i. More specifically, the function $p(x|H_i)$ is the probability density that the OBS operation outputs x for phenomena satisfying hypothesis H_i. The densities are normalized to unity, $\int_{\mathcal{D}} p(x|H_i)\,dx = 1$, where $\mathcal{D} \subseteq \mathcal{R}$ is the region of allowed x-values. The adaptive system is trained on the sets $\{x_1^1,\ldots,x_{M_1}^1\},\ldots,\{x_1^N,\ldots,x_{M_N}^N\}$ of OBS data outputs for each hypothesis $H_1,\ldots,H_N$. This training set results from M_1 trials of OBS with hypothesis H_1, M_2 trials of OBS with hypothesis H_2, continuing to M_N trials of OBS with hypothesis H_N. The system is trained to exactly perform the mapping

$$x_i^j \longmapsto (0,\ldots,0,\overbrace{1}^{j},0,\ldots,0),\ i = 1,\ldots,M_j,\ j = 1,\ldots,N \quad . \tag{1}$$

A measure of system errors due to inherent training-set ambiguities is obtained from the performance on the training set $\{x_1^1,\ldots,x_{M_1}^1\},\ldots,\{x_1^N,\ldots,x_{M_N}^N\}$. This intuitively represents an upper bound on averaged system performance because added errors generally occur due to incorrect system association on arbitrary data. To compute the training-set-based measures, it is assumed that $M_1+\cdots+M_N$ trials of OBS result exactly in the data set $\{x_1^1,\ldots,x_{M_1}^1\}\cup\cdots\cup\{x_1^N,\ldots,x_{M_N}^N\}$. For a given data point $x_i^j, i = 1,\ldots,M_j, j = 1,\ldots,N$, the probability of having been generated by hypothesis $H_k, k = 1,\ldots,N$, is given by

$$Prob(x_i^j, H_k) = \frac{p(H_k)p(x_i^j|H_k)}{\sum_{q=1}^N p(H_q)p(x_i^j|H_q)} \quad , \tag{2}$$

where the normalization is over the hypotheses which could have generated x_i^j in the $M_1+\cdots+M_N$ trials. The system maps x_i^j to hypothesis H_j so that the probability in Equation (2) contributes to the situation of a system declaration for hypothesis H_j when the true hypothesis is H_k. Therefore, over the set of $M_1+\cdots+M_N$ trials of OBS, the number of H_j declarations for true hypothesis H_k is given from Equation (2) by

$$\begin{aligned} NUM(H_j,H_k) &= \sum_{i=1}^{M_j} Prob(x_i^j, H_k) \\ &= \sum_{i=1}^{M_j} \frac{p(H_k)p(x_i^j|H_k)}{\sum_{q=1}^N p(H_q)p(x_i^j|H_q)} \quad . \end{aligned} \tag{3}$$

The probability of a system declaration of H_j for true hypothesis H_k is then given by $(j,k = 1,\ldots,N)$,

$$p(H_j, H_k) = \frac{1}{\sum_{p=1}^{N} M_p} \sum_{i=1}^{M_j} \frac{p(H_k)p(x_i^j|H_k)}{\sum_{q=1}^{N} p(H_q)p(x_i^j|H_q)} \quad . \tag{4}$$

Note that the required normalization for the $M_1 + \cdots + M_N$ trials, $\sum_{k=1}^{N} p(H_j, H_k) = M_j / \sum_{p=1}^{N} M_p$, follows from Equation (4). We now consider the average of $p(H_j, H_k)$ over the ensemble of training sets obtained by the above procedure. Recall that x_i^j in Equation (4) was obtained by the OBS operation with a fixed hypothesis H_j, indicating that the appropriate distribution for x_i^j is $p(x_i^j|H_j)$. Averaging over the values of x_i^j in Equation (4), we obtain an averaged probability for hypothesis H_j declared with the true hypothesis H_k,

$$\langle p(H_j, H_k) \rangle = \gamma_j p(H_k) \rho_{j,k}, \; j, k = 1, \ldots, N \quad , \tag{5}$$

where

$$\rho_{j,k} = \int_{\mathcal{D}} \frac{p(x|H_j)p(x|H_k)}{\sum_{q=1}^{N} p(H_q)p(x|H_q)} \, dx \tag{6}$$

and γ_j is the proportion of hypothesis H_j-generated data in the training set for the adaptive system

$$\gamma_j = \frac{M_j}{\sum_{q=1}^{N} M_q} \quad . \tag{7}$$

The joint probability in Equation (5) has factored into a training-ensemble-dependent parameter γ_j and a statistics-dependent quantity $p(H_k)\rho_{j,k}$. An estimate of the conditional probability $p(H_i|H_j)$, corresponding to a decision for H_i with true hypothesis H_j, is obtained from Equation (5) by

$$\begin{aligned} p(H_i|H_j) &= \frac{\langle p(H_i, H_j) \rangle}{\sum_{q=1}^{N} \langle p(H_q, H_j) \rangle} \\ &= \frac{\gamma_i \rho_{i,j}}{\sum_{q=1}^{N} \gamma_q \rho_{q,j}} \quad , \end{aligned} \tag{8}$$

where γ_i and $\rho_{i,j}$ are given in Equations (5) and (6), respectively. Equations (5)–(8) are denoted the training-set-based measures of system performance in the following sections.

2.2 Neural-Net MAP Test Measures

A more traditional approach to system performance estimation is through the MAP test [2]. For an OBS-generated input x, the hypothesis H_j is chosen, which maximizes the conditional probability $p(H_k|x)$, $k = 1, \ldots, N$. A neural net trained on sufficiently representative data has been

shown to converge to the MAP test performance [13]. A mapping network for the N-hypothesis test consists of a single OBS-generated input x, a series of hidden layers, and an N-neuron output layer. A stochastic formulation of a MAP test neural net allows a comparison with the training-set-based estimates in Equations (5)–(8). The N deepest layer neurons are assumed to output only 0 or 1 in the pattern $(0,\ldots,0,\overbrace{1}^{i},0,\ldots,0), i=1,\ldots,N$, with probability $q_i(x)$ for input x. The output N-vector $(0,\ldots,0,\overbrace{1}^{i},0,\ldots,0)$ corresponds to a decision for hypothesis H_i. The net output probabilities are normalized by the condition $\sum_{j=1}^{N} q_j(x) = 1$, $x\epsilon\mathcal{D}$. The joint probability $p(H_j, H_k|x)$ for choosing hypothesis H_j with phenomena satisfying H_k, assuming net input x, is given by the product $q_j(x)p(H_k|x)$. The average over input values x with a prior distribution $p(x)$ yields

$$p^{MAP1}(H_j, H_k) = \int_{\mathcal{R}} q_j(x)p(H_k|x)p(x)\,dx \quad . \tag{9}$$

The MAP test follows on average for

$$q_j(x) = \frac{p(H_j|x)}{\sum_{q=1}^{N} p(H_q|x)} \quad . \tag{10}$$

Substitution of Equation (10) into Equation (9) yields, upon application of Bayes's theorem,

$$p(H_j|x) = \frac{p(x|H_j)p(H_j)}{p(x)},\ j=1,\ldots,N \quad , \tag{11}$$

the equation

$$p^{MAP1}(H_j, H_k) = p(H_j)p(H_k)\rho_{j,k},\ j,k=1,\ldots,N \quad , \tag{12}$$

where $\rho_{j,k}$ is defined in Equation (6). Comparison of Equations (5) and (12) suggests that the MAP test estimate equals the training-set-based estimate if the training set satisfies the equation $\gamma_j = p(H_j)$. This condition reflects the common-sense belief that the training set should be proportioned according to the prior probabilities of the hypotheses.

A deterministic neural-net model for the MAP N-hypothesis test occurs if the N-deepest layer neurons output analog values in the range [0,1]. We assume that for net input $x\epsilon\mathcal{D}$ the ith, $i=1,\ldots,N$, neuron literally outputs the value $p(H_i|x)$. The MAP test then results simply from choosing the hypothesis H_i corresponding to the deepest layer neuron with the largest output value. A schematic of the deterministic MAP test neural net is shown in Figure 2. In order to

compute performance probabilities for this net, we must define regions $\mathcal{B}_j$, $j = 1, \ldots, N$, given

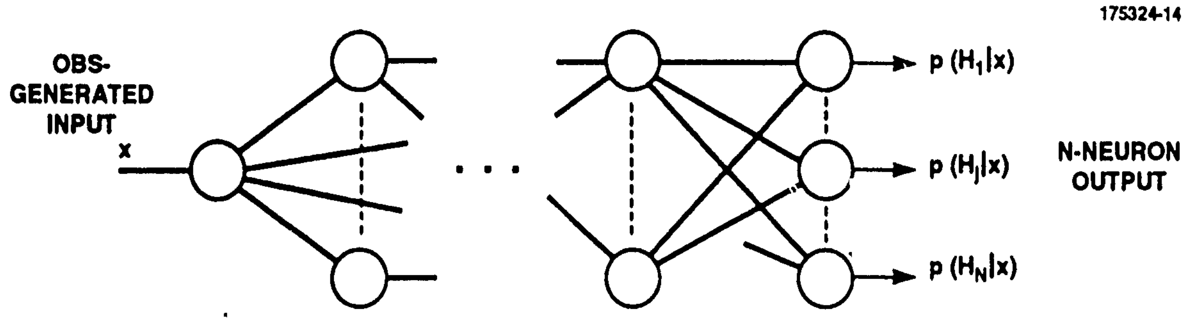

Figure 2. Schematic of deterministic neural-net representation of MAP test. OBS-generated input x, N-hypothesis neuron output.

by $\mathcal{B}_j = \{x \epsilon \mathcal{D} | p(H_j|x) > p(H_k|x), \forall k \neq j\}$. Assuming that the regions of equal conditional probabilities, $\mathcal{E}_{j,k} = \{x \epsilon \mathcal{D} | p(H_j|x) = p(H_k|x)\}$, $j, k = 1, \ldots, N$, have zero support, we define the joint performance probability $p^{MAP2}(H_j, H_k)$ by the expression

$$p^{MAP2}(H_j, H_k) = p(H_k) \int_{\mathcal{B}_j} p(x|H_k)\, dx \qquad , \tag{13}$$

corresponding to the probability that the jth neuron output in Figure 2 is maximum for an H_k-generated input. The computation of the performance probabilities $p^{MAP2}(H_j, H_k)$, $j, k = 1, \ldots, N$ follows from the application of Bayes's formula in Equation (11) to the definitions of regions $\mathcal{B}_j$ and $\mathcal{E}_{j,k}$. In Section 3, Equation (13) is applied to the binary hypothesis test on uniformly distributed data for comparison with the training-set-based estimates in Equations (5)–(8).

3. BINARY HYPOTHESIS TEST

In this section performance measures derived in Section 2 are applied to a binary hypothesis test of uniformly distributed data. The averaged probabilities in Equations (8) and (13) are related to parameters in the hypothesis probability distributions. This relationship defines a framework for comparison of the estimates with examples of adaptive system performance.

3.1 Training-Set-Based and MAP Estimates

Consider a training-set-based decision between hypothesis H_0 and H_1 with prior probabilities $p_0 = p(H_0)$ and $p_1 = p(H_1)$, respectively. Assume the output from the OBS operation, x, has conditional probabilities $p(x|H_i)$, $i = 0, 1$, for phenomena satisfying hypothesis H_i. The system performance is defined by the standard conditional probabilities of detection $P_d = p(H_1|H_1)$, false alarm $P_f = p(H_1|H_0)$, miss $P_m = p(H_0|H_1)$, and the correct H_0 identification $P_{cH_0} = p(H_0|H_0)$. Assuming a training set consisting of N_i, $i = 0, 1$, trials of OBS with hypothesis H_i, we have, from Equation (8),

$$P_d = \frac{\gamma_1 \rho_{1,1}}{\gamma_1 \rho_{1,1} + \gamma_0 \rho_{0,1}} \quad , \tag{14}$$

$$P_f = \frac{\gamma_1 \rho_{1,0}}{\gamma_1 \rho_{1,0} + \gamma_0 \rho_{0,0}} \quad , \tag{15}$$

$$P_m = \frac{\gamma_0 \rho_{0,1}}{\gamma_0 \rho_{0,1} + \gamma_1 \rho_{1,1}} \quad , \tag{16}$$

and

$$P_{cH_0} = \frac{\gamma_0 \rho_{0,0}}{\gamma_0 \rho_{0,0} + \gamma_1 \rho_{1,0}} \quad , \tag{17}$$

where $\gamma_i = N_i/(N_0 + N_1)$, $i = 0, 1$, and

$$\rho_{j,k} = \int_{\mathcal{D}} \frac{p(x|H_j)p(x|H_k)}{p_0 p(x|H_0) + p_1 p(x|H_1)}\, dx\,, \quad j,k = 0,1 \quad , \tag{18}$$

with $\mathcal{D}$ the region of possible x-values.

A common situation that occurs in the conventional Neyman Pearson test is the existence of a maximum tolerated joint false alarm probability $P_F = p(H_1, H_0)$. From Equation (5), a maximum joint false alarm probability P_{F_0} implies an upper bound on the percentage of H_1 trials in the training set, i.e., the condition $\gamma_1 < P_{F_0}/p_0\rho_{1,0}$. There is also a corresponding upper bound on the joint detection probability $P_D = p(H_1, H_1)$, given by $P_D < P_{F_0} p_1 \rho_{1,1}/p_1 \rho_{1,0}$.

The MAP test performance measure in Equation (13) can also be applied to the binary hypothesis test. Assume that for $x \epsilon \mathcal{E}_{0,1}$, which is the region of equal *a posteriori* probability, the test chooses between hypothesis H_0 and H_1 with equal probability. In this case the neural net has equal output values from the two deepest layer neurons in Figure 2. The expression in Equation (13) is easily generalized to obtain the conditional probabilities

$$P_d^{MAP2} = \int_{\mathcal{B}_1} p(x|H_1)\,dx + \frac{1}{2}\int_{\mathcal{E}_{0,1}} p(x|H_1)\,dx \quad , \tag{19}$$

$$P_f^{MAP2} = \int_{\mathcal{B}_1} p(x|H_0)\,dx + \frac{1}{2}\int_{\mathcal{E}_{0,1}} p(x|H_0)\,dx \quad , \tag{20}$$

$$P_m^{MAP2} = \int_{\mathcal{B}_0} p(x|H_1)\,dx + \frac{1}{2}\int_{\mathcal{E}_{0,1}} p(x|H_1)\,dx \quad , \tag{21}$$

and

$$P_{cH_0}^{MAP2} = \int_{\mathcal{B}_0} p(x|H_0)\,dx + \frac{1}{2}\int_{\mathcal{E}_{0,1}} p(x|H_0)\,dx \quad . \tag{22}$$

In order to compare performance measures in Equations (14)-(17) with the MAP estimates in Equations (19)-(22), consider the case of uniformly distributed conditional probabilities $p(x|H_i)$ of width Δ_i, $i = 0,1$. Figure 3 contains uniform distributions $p(x|H_i)$, $i = 0,1$, normalized to a peak value of $1/\Delta_i$, centered at 0 for H_0 and at x_1 for H_1. The distributions are overlapped under the condition $|\Delta_0 - \Delta_1|/2 \leq x_1 \leq (\Delta_0 + \Delta_1)/2$. Substitution of the uniform distributions in Equation (18) results in expressions for $\rho_{i,j}$, $i,j = 0,1$, given by

$$\rho_{0,0} = \frac{(x_1 + \frac{\Delta_0 - \Delta_1}{2})}{p_0\Delta_0} + \frac{\Delta_1(\frac{\Delta_0+\Delta_1}{2} - x_1)}{\Delta_0(p_0\Delta_1 + p_1\Delta_0)} \quad , \tag{23}$$

$$\rho_{0,1} = \rho_{1,0} = \frac{(\frac{\Delta_0+\Delta_1}{2} - x_1)}{p_0\Delta_1 + p_1\Delta_0} \quad , \tag{24}$$

and

$$\rho_{1,1} = \frac{(x_1 + \frac{\Delta_1 - \Delta_0}{2})}{p_1\Delta_1} + \frac{\Delta_0(\frac{\Delta_0+\Delta_1}{2} - x_1)}{\Delta_1(p_0\Delta_1 + p_1\Delta_0)} \quad . \tag{25}$$

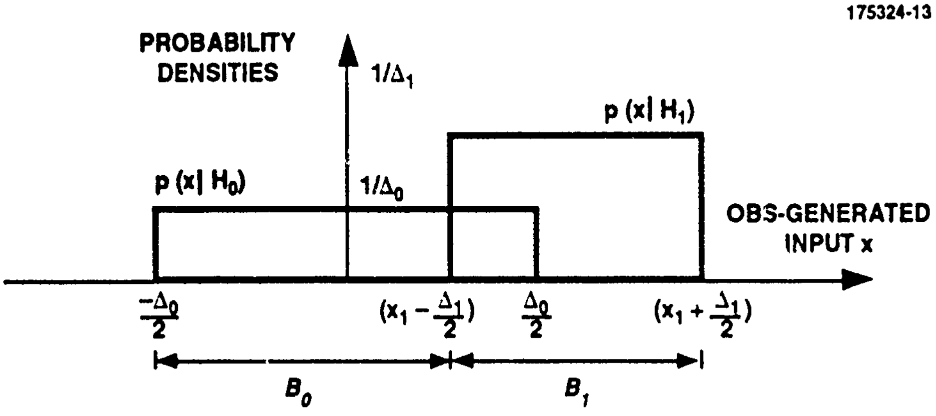

Figure 3. Binary hypothesis test on uniformly distributed data. Input probability distributions $p(x|H_i)$, $i = 0, 1$. Width Δ_i, $i = 0, 1$; center 0 for H_0 and x_1 for H_1.

The adaptive system simulation in Section 3.2 is for uniform probability distributions of equal width, $\Delta_0 = \Delta_1 = \Delta$, separated by $x_1 = K\Delta$. The K-factor parameterization of overlapped distributions is convenient for analysis of system discrimination performance [15]. The overlapped distribution condition corresponds to $K\epsilon[0, 1]$, with K of unity for non-overlapped distributions. The training-set-based measures for uniform distributions are obtained from the substitution of Equations (23)–(25) into Equations (14)–(18), with the result

$$P_d = \frac{\gamma_1[K + (1-K)p_1]}{\gamma_1 K + (1-K)p_1} \quad , \tag{26}$$

$$P_f = \frac{\gamma_1 p_0 (1-K)}{\gamma_0 K + (1-K)p_0} \quad , \tag{27}$$

$$P_m = \frac{\gamma_0 p_1 (1-K)}{\gamma_1 K + (1-K)p_1} \quad , \tag{28}$$

and

$$P_{cH_0} = \frac{\gamma_0[K + (1-K)p_0]}{\gamma_0 K + (1-K)p_0} \quad . \tag{29}$$

The MAP test measures in Equations (19)–(22) can be computed for the uniform data distributions in Figure 3. Note that for the case $\Delta_0 = \Delta_1 = \Delta$, $x_1 = K\Delta$, and $p_0 = p_1 = 0.5$, the region of equal *a posteriori* probability $\mathcal{E}_{0,1}$ is $[\Delta(K - \frac{1}{2}), \Delta/2]$; the dominant hypothesis regions are given by $\mathcal{B}_0 = [-\Delta/2, \Delta(K - \frac{1}{2})]$ and $\mathcal{B}_1 = [\Delta/2, \Delta(K + \frac{1}{2})]$. Substitution of these regions into Equations (19)–(22) with uniform conditional probabilities $p(x|H_i)$, $i = 0, 1$, yields

$$P_d^{MAP2} = P_{cH_0}^{MAP2} = (1 + K)/2 \tag{30}$$

and

$$P_m^{MAP2} = P_f^{MAP2} = (1 - K)/2 \quad . \tag{31}$$

Note that for the case $\gamma_0 = p_0 = 0.5$ and $\gamma_1 = p_1 = 0.5$, Equations (26)–(29) and Equations (30) and (31) are identical, as expected for a training set proportioned according to prior probabilities. Note that the condition $p_0 = p_1$ implies that the MAP test is equivalent to the maximum likelihood test, which maximizes $p(x|H_j)$, $j = 1, \ldots, N$, to determine the hypothesis.

3.2 Performance Estimate Comparisons

In the analysis of the previous sections, performance measures for an adaptive system were obtained from the statistics of the training set. We also derived estimates based on the assumption that an adaptive system, realized as a neural net, performs a MAP N-hypothesis test. Back-propagation neural nets are adaptive systems consisting of connected layers of processing elements (neurons) with adjustable connection weights between layers and adjustable thresholds on each neuron. A training set for decision making is used to adjust net parameters so that the net performs a map between the input data space and the output hypothesis space. Both the training-set-based and MAP estimates defined in the previous sections involve particular assumptions about the adaptive system. The training-set-based estimate assumes that the system power of association, i.e., the ability to decide on data not trained on, does not affect the system performance. The MAP estimate for neural nets assumes that regardless of training-set composition the network literally outputs the conditional probabilities $p(H_i|x)$, $i = 1, \ldots, N$, in the N-neuron deepest layer. In this section we compare the training-set-based- and MAP-performance measures on the binary hypothesis test in Section 3.1. The performance of two adaptive systems, a linear extrapolation from the data set (LINEXT) and a back-propagation neural net (BPNN), are compared with the two performance measures.

Figure 4 contains plots of $P_d(K)$ and $P_f(K)$ for the training-set-based estimate from Equations (26)–(29) and the MAP estimate from Equations (30) and (31) for the binary test of uniformly distributed data described in Section 3.1. We assume equal prior probabilities for H_0 and

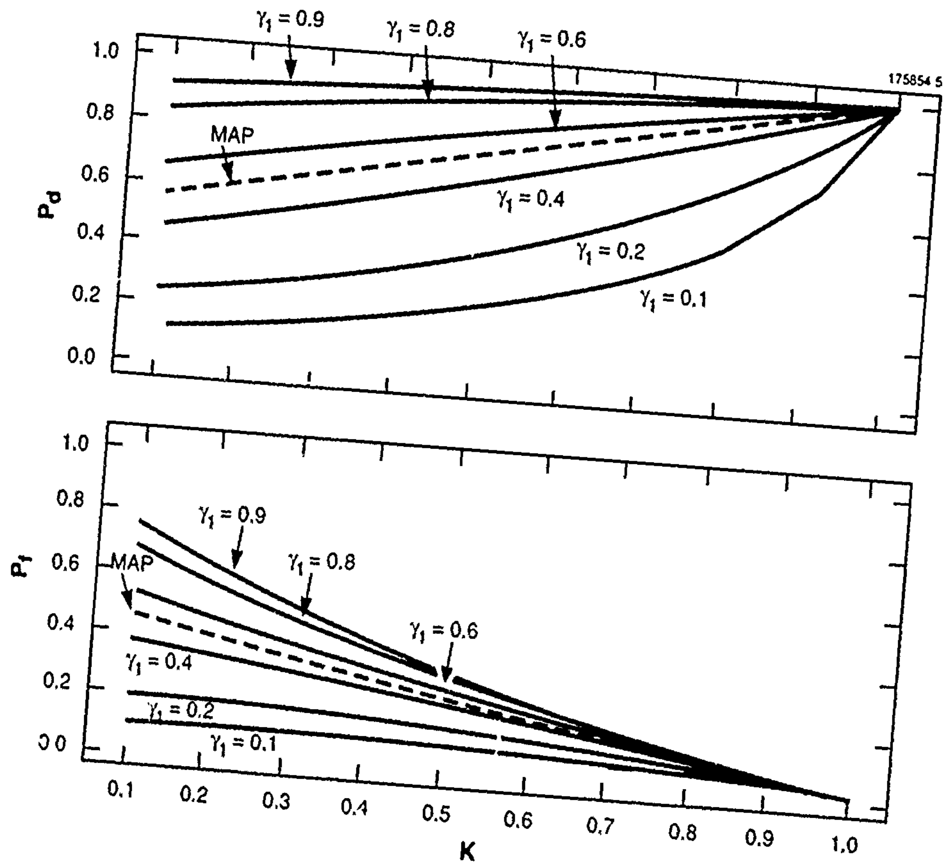

Figure 4. Detection and false alarm probability versus K for binary hypothesis test. Training-set-based estimate with $\gamma_1 = 0.1, 0.2, 0.4, 0.6, 0.8, 0.9$ and MAP estimate. Prior probabilities $p_0 = 0.5$, $p_1 = 0.5$

H_1, $p_0 = p_1 = 0.5$, and plot the conditional probabilities for various values of γ_1. Note that if half

the training set is H_1-generated, $\gamma_1 = 0.5$, the training-set-based probabilities are linear in K and match the MAP estimates. The results in Figure 4 indicate that a training set proportioned toward H_1, i.e., $\gamma_1 > \gamma_0$, increases P_d (at the expense of P_f) over the MAP test estimate. The reverse situation occurs for a training set proportioned toward H_0.

A simple adaptive algorithm for hypothesis testing on binary hypotheses is linear extension (LINEXT) from the training set. Assume a training set of OBS-generated data $\{x_1^0, \ldots, x_{N_0}^0\} \cup \{x_1^1, \ldots, x_{N_1}^1\}$, where x_j^i is H_i-generated. An input x is mapped by LINEXT to the hypothesis of the nearest element of the training set. If the adaptive system is viewed as a map f from $\mathcal{D}$ to $\{0,1\}$, with $f(x) = i$ for hypothesis H_i, then the nearest-training-set-neighbor algorithm is simply a linear extension of f from the training set. The hypothesis chosen for input x results from a decision threshold at 0.5, e.g., $f(x)$ greater (less) than 0.5 implies hypothesis H_1 (H_0). The performance of the LINEXT algorithm was tested by creating a training set with $N_0 = 100\gamma_0$ and $N_1 = 100\gamma_1$ H_0 and H_1-generated elements, respectively. We considered the cases of $\gamma_0 = 0.1$ and 0.2 separately and in each case used a training ensemble consisting of 1000 training sets. The LINEXT algorithm for each training set was tested with an independent performance set of 400 elements. Each performance-set element was generated by first choosing H_0 or H_1 phenomena according to p_0 and p_1 prior probabilities. The chosen hypothesis H_i determined the distribution $p(x|H_i)$ (Figure 3) used to generate the data value $x \epsilon \mathcal{D}$. The LINEXT algorithm was applied to $x \epsilon \mathcal{D}$ and the output hypothesis H_j was compared to the originating hypothesis H_i. The number of elements mapped to H_j originating from an observation of H_i, divided by the number of elements in the performance set from H_i, yielded the conditional probability $p(H_j|H_i)$. Figures 5 and 6 show the average performance of the LINEXT algorithm in which the performance set probability estimates described above were averaged over the training ensemble of 1000 sets. Figure 5 contains a plot of P_d, P_f, P_m, and P_{cH_0} as a function of K for an ensemble of training sets with $\gamma_0 = 0.2$ and $\gamma_1 = 0.8$. Note that the experimental performance of LINEXT was well approximated by the training-set-based estimate (dashed line), particularly for P_d and P_m probabilities. As seen in Figure 6, similar results were obtained for an ensemble of training sets with $\gamma_0 = 0.1$ and $\gamma_1 = 0.9$. Note that the MAP test estimate (dotted line) provided a less successful prediction of LINEXT performance.

The LINEXT algorithm above performs the decision space mapping on the training-set elements exactly. For a sufficiently representative training set in the overlap region of Figure 3, this necessitates a mapping with undulations between the H_0 and H_1 hypotheses. A three-layer BPNN has proven sufficient to perform any reasonable functional mapping [16]. A rough estimate of the required number of neurons is obtained from the BPNN threshold function,

$$T_\theta(I) = \frac{1}{1 + \exp(-I + \theta)} \quad , \tag{32}$$

which is applied to the input I of a neuron with threshold value θ. An undulation in the net input/output mapping is easily represented as the difference of two neuron threshold functions,

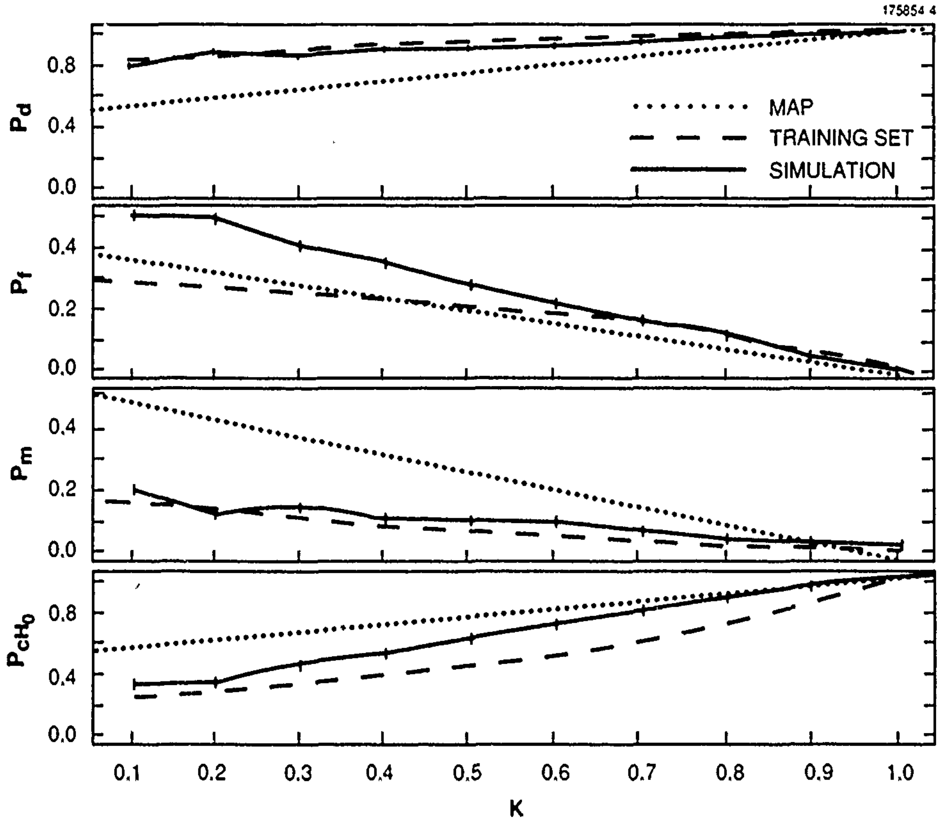

Figure 5. Detection, false alarm, miss, and correct H_0 probabilities versus K for LINEXT algorithm on binary hypothesis test. Averaged over 1000 training sets with $\gamma_0 = 0.2$ and $\gamma_1 = 0.8$. Prior probabilities $p_0 = p_1 = 0.5$. Each trained system performance-tested with 400 elements.

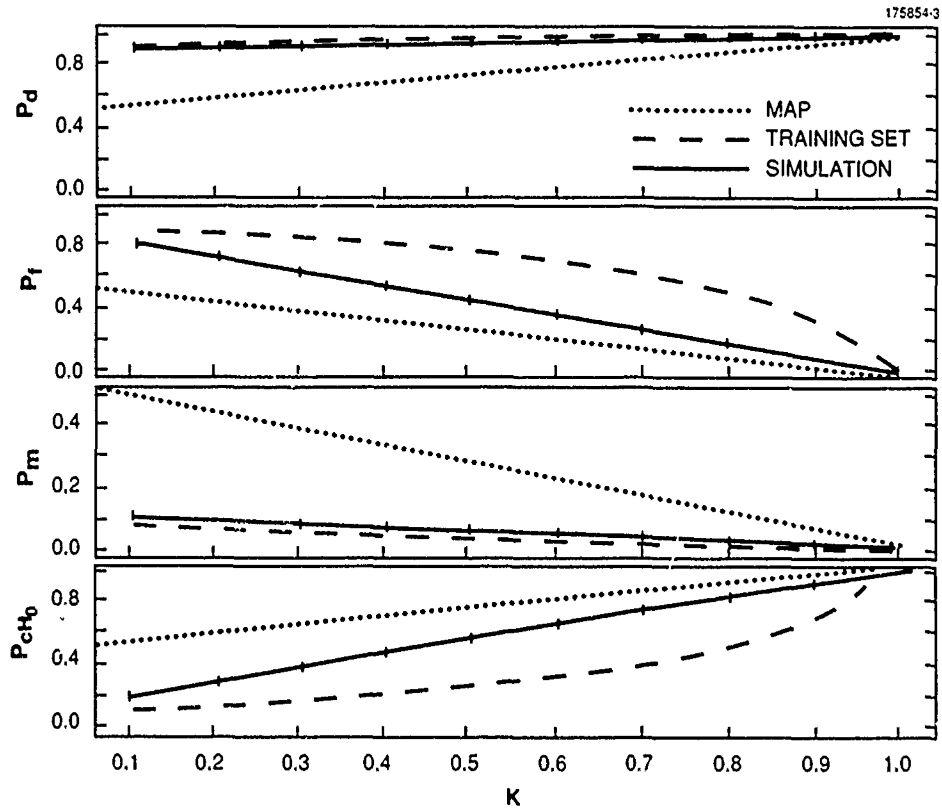

Figure 6. Detection, false alarm, miss, and correct H_0 probabilities versus K for LINEXT algorithm on binary hypothesis test. Averaged over 1000 training sets with $\gamma_0 = 0.1$ and $\gamma_1 = 0.9$. Prior probabilities $p_0 = p_1 = 0.5$. Each trained system performance-tested with 400 elements.

$T_\theta - T_{\theta'}$. This suggests that a mapping with p undulations requires at least $2p$ hidden-layer neurons in the three-layer net. However, a BPNN trained to exactly perform the hypothesis space map over a training set most likely approximates the training-set-based performance estimate. A neural net with performance matching the MAP estimate is trained on data biases rather than on an exact training-set map. In order to obtain MAP test performance, we considered a BPNN with a single input, sixteen hidden-layer neurons, and two output neurons. The net was trained to perform the binary hypothesis test on uniformly distributed data of equal width ($\Delta_0 = \Delta_1 = \Delta$) and equal prior probabilities ($p_0 = p_1 = 0.5$). For each training set of 20 H_0- and 80 H_1-generated inputs ($\gamma_0 = 0.2, \gamma_1 = 0.8$), the net was trained to map to (1,0) and (0,1) for H_0 and H_1, respectively. To avoid mapping to training-set undulations and to train only on data biases, the inputs from the overlapped regions in Figure 3 were removed from the training sets. The performance probabilities for the trained BPNNs as a function of K are shown in Figure 7. For each K, ten BPNNs were trained on independent training sets of 20 and 80 H_0- and H_1-generated inputs. For each trained BPNN, a set of 1000 four-hundred-element performance sets, each with equal prior probabilities p_i of 0.5, was used to compute performance probabilities. The BPNN output decisions were determined by the larger neuron outputs in the third layer. As with the LINEXT algorithm, the conditional probabilities $p(H_i|H_j)$ were determined by counting the number of H_i decisions from H_j-generated data and dividing by the total number of H_j-generated elements in the performance set. As demonstrated in Figure 7, although the training set was proportioned toward H_1 ($\gamma_1 = 0.8$) the BPNN performance was best approximated by the MAP estimate (dotted line). These results highlight the fundamental difference between training-set-based- and MAP-test estimation of adaptive system performance.

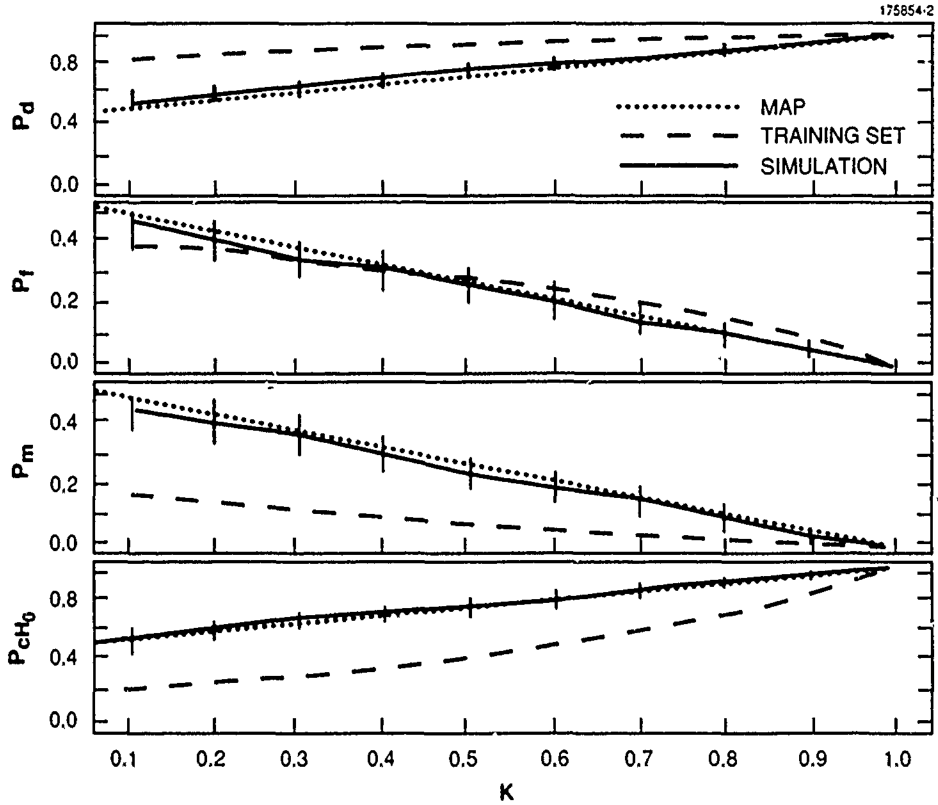

Figure 7. Detection, false alarm, miss, and correct H_0 probabilities versus K for BPNN algorithm on the binary hypothesis test. Averaged over 10 training sets with $\gamma_0 = 0.2$ and $\gamma_1 = 0.8$. Prior probabilities $p_0 = p_1 = 0.5$. Each trained system performance-tested with 1000 sets of 400 elements each.

4. CONCLUSION

In this report two distinct performance measures were identified for adaptive systems. Training-set-based estimation of system performance was derived from the statistics of the training set. These statistics are relevant if system errors reflect uncertainties inherent in the learning procedure. The measures are independent of a particular adaptive system, although it was argued that systems which perform training-set map undulations are described by training-set-based estimates. The training-set-based measures were compared to the performance of a MAP test, which is easily represented in a neural net. It was suggested that systems trained for data biases rather than an exact training-set map are best described by the MAP test performance.

Two adaptive systems were considered to emphasize the differences between training-set-based- and MAP-test performance measures. The LINEXT algorithm, as applied to the binary hypothesis test, performed the training-set map exactly. It was experimentally determined that the LINEXT system performance was well approximated by the training-set-based estimate. Alternatively, it was shown that a BPNN trained on data biases had a performance matching the MAP test estimate.

The desired system performance has implications for neural-net structure. For example, it was argued that two neurons are required in a three-layer BPNN for each implemented undulation in the training-set map. An adaptive system matching MAP test performance would not have this structural condition. However, training-set-based performance may be desirable because performance probabilities are dependent on the training set. For example, a training set proportioned toward particular hypotheses increases the system performance for conditional probabilities involving those hypotheses. In this report a Neyman-Pearson-like bound for the binary hypothesis test was shown to imply an upper bound on γ_1, the proportion of detection data in the training set; the adaptive system must be described by the training-set-based estimate for such bounds to be relevant.

REFERENCES

1. R. Hecht-Nielson, *Neurocomputing*, Reading, MA: Addison Wesley (1990).
2. H. Van Trees, *Detection, Estimation, and Modulation Theory, Part I*, New York: John Wiley and Sons (1971).
3. R.P. Lippmann, "An introduction to computing with neural nets," *IEEE ASSP Magazine* 4, 4–22 (1987).
4. D.E. Rumelhart, G.E. Hinton, and R.J. Williams, "Learning internal representation by error propagation," in D.E. Rumelhart and J.L. McClelland (eds.), *Parallel Distributed Processing: Explorations in the Microstructure of Cognition 1*, Cambridge, MA: MIT Press (1986), pp. 318-362.
5. P.J. Werbos, doctoral dissertation, Harvard University, Cambridge, MA (1974).
6. D.B. Parker (memo, 1985).
7. A.E. Bryson and Y.C. Ho, *Applied Optimal Control* (Revised Second Printing), New York: Hemisphere Publishing (1975).
8. T. Kohonen, *Self-Organization and Associative Memory*, ed. 2, Berlin: Springer-Verlag (1988).
9. W.Y. Huang and R.P. Lippmann, "Comparisons between neural-net and traditional classifiers," in *Proc. of IEEE First Int. Conf. on Neural Nets*, San Diego, CA., June, 1987 (IEEE Press, New York, 1987), pp. IV485–IV493.
10. W.Y. Huang and R.P. Lippmann, "Neural-net and traditional classifiers," in D. Anderson (ed.), *Neural Information Processing Systems*, New York: American Institute of Physics (1988), pp. 387–396.
11. Murphy, O.J., "Nearest neighbor pattern classification perceptrons," *Proc. IEEE* 78, 1595-1598 (1990).
12 H-C. Yau and M.T. Manry, "Iterative improvement of a Gaussian classifier," *Neural Networks* 3, 437–443 (1990).
13. Y. Yair and A. Gersho, "Maximum *a posteriori* decision and evaluation of class probabilities by Boltzmann perceptron classifiers," *Proc. IEEE* 78, 1620–1628 (1990).
14. L.I. Perlovsky and M.M. McManus, "Maximum likelyhood neural nets for sensor fusion and adaptive classification," *Neural Networks* 4, 89–102 (1991).
15. K. Fukunaga, *Introduction to Statistical Pattern Recognition*, New York: Academic Press (1972).
16. R. Hecht-Nielson, "Theory of back-propagation neural nets," *Proc. of IEEE First Int. Conf. on Neural Nets*, San Diego, CA., June, 1987 (IEEE Press, New York, 1987), pp. I593–I611.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY (*Leave blank*)

2. REPORT DATE
11 September 1991

3. REPORT TYPE AND DATES COVERED
Technical Report

4. TITLE AND SUBTITLE
Performance Measures for Adaptive Decisioning Systems

5. FUNDING NUMBERS
C — F19628-90-0002
PE — 12424F, 31310F, 63223C
PR — 80

6. AUTHOR(S)
Robert Y. Levine
Timothy S. Khuon

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Lincoln Laboratory, MIT
P.O. Box 73
Lexington, MA 02173-9108

8. PERFORMING ORGANIZATION REPORT NUMBER
TR-927

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
U.S. Army Strategic Defense Command
P.O. Box 1500
Huntsville, AL 35807-3801

10. SPONSORING/MONITORING AGENCY REPORT NUMBER
ESD-TR-91-094

11. SUPPLEMENTARY NOTES
None

12a. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release: distribution is unlimited.

12b. DISTRIBUTION CODE

13. ABSTRACT (*Maximum 200 words*)

Performance measures are derived for data-adaptive hypothesis testing by systems trained on stochastic data. The measures consist of the averaged performance of the systems over an ensemble of training sets. The uncertainties derivable from training sets represent an irreducible uncertainty inherent in the learning procedure. Data-adaptive system estimates are contrasted with classical hypothesis testing, in which optimum tests are based on an assumed data model. In addition, a performance estimate for the maximum *a posteriori* probability (MAP) N-hypothesis test is derived based on a neural-net formulation of the test. The performance of adaptive systems on a binary test of uniformly distributed data is compared with the data-adaptive and MAP estimates. The adaptive systems considered are linear extrapolation from data (LIN-EXT) and a back-propagation neural net (BPNN).

14. SUBJECT TERMS
adaptive decisioning systems; performance estimation; neural nets; maximum *a posteriori* probability (MAP)

15. NUMBER OF PAGES
34

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	SAR

NSN 7540-01-280-5500

Standard Form 298 (Rev. 2-89)
Prescribed by AMSI Std. 239-18
298-102